Objective Prediction of Speech Quality for Wideband Communication Scenarios including background noise

H.W. Gierlich, F. Kettler, S. Poschen, J. Reimes

HEAD acoustics GmbH, Herzogenrath

1. Introduction

Speech quality in presence of background noise is of great importance in today's communication systems. Mobile terminals but also hands-free devices are used more and more in noisy environments. Known and standardized speech quality measures like PESQ and TOSQA2001 are not suitable for those scenarios. The speech quality in presences of background noises can be separated in three components:

- quality of transmitted background noise
- quality of transmitted speech
- overall quality.

According to ITU-T Recommendation P.835 these parameters can be assessed in subjective listening tests. Within the ETSI / eEurope STF 294 project a database was generated in order to develop a model capable to predict speech, noise and overall quality in wideband scenarios.

2. Database

The database for the listening test and the objective model contains speech samples in two languages, a female and a male speaker each, recordings of wideband handsets and hands-free devices, different background noises (car, office, pub, …), different wideband speech coders, several noise reduction and VAD implementations and several network conditions. In the listening tests the speech, noise and overall quality was rated on a five point ACR scale [1].

3. Relative Approach

The new objective model is based on Relative Approach [3], an aurally adequate analysis. The Relative Approach models the characteristic of the human hearing to perceive much stronger patterns (tones, relatively rapidly time-varying structures) than slowly changing levels and loudness. The Relative Approach analysis is based on the assumption that human hearing continuously creates a reference sound (an "anchor signal") for its automatic recognition process against which it classifies tonal or temporal pattern information moment-by-moment. It evaluates the difference between the instantaneous pattern in both time and frequency and the "smooth" or less-structured content in similar time and frequency ranges. In evaluating the acoustic quality of a complex "patterned" signal, the absolute level or loudness is almost without any significance. Temporal structures and spectral patterns are important factors in deciding whether a sound is judged as annoying or disturbing [2], [3], [4].

As the human hearing the Relative Approach does not require a reference signal. Comparable to the human experience and expectation, the algorithm generates an "internal reference" which can be best described as a forward estimation. The Relative Approach algorithm objectivizes pattern(s) in accordance with human perception by resolving or extracting them while largely rejecting pseudostationary energy. At the same time, it considers the context of the relative difference of the "patterned" and "non-patterned" magnitudes. Here a variant of Relative Approach is used which analyses the *changes* of speech and noise sound before and after the transmission. This leads to the Δ Relative Approach.

4. Objective N-/S-/G-MOS

4.1 Calculation of N-MOS

The following parameters were found to mainly influence the background noise "quality": the absolute background noise level, the modulation of background noise, e.g. musical tones, "naturalness" and lost packets influence).

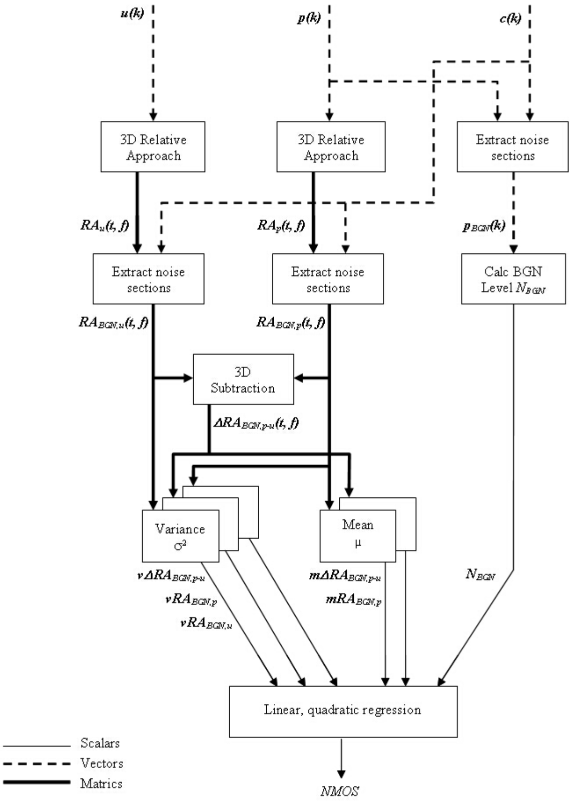

Figure 1: Block diagram of N-MOS calculation

The principle of the N-MOS calculation is shown in figure 1. The N-MOS algorithm focuses on the characterization of the changes in the background noise due to the transmission. For this characterization the parameters mean and variance are calculated based on the 3D Δ Relative Approach between the transmitted signal p(k) and the non-transmitted signal + noise u(k). By the mean and the variance the "similarity" of the non-transmitted and the transmitted background noise is determined. Together with the absolute level of the transmitted background noise they are fed to a linear quadratic regression leading to the objective N-MOS. The correlation of the N-MOS algorithm to the auditory N-MOS is described in the following.

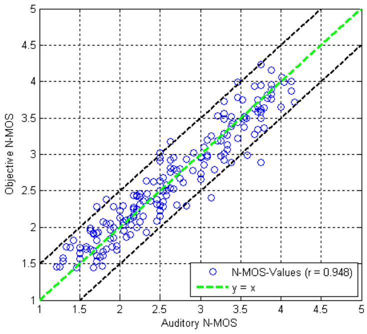

Figure 2: Obj. calculated N-MOS versus audit. N-MOS

Figure 2 shows the objectively calculated N-MOS vs. the auditory N-MOS. The per sample deviation between the subjective and objective N-MOS is less than 0.5 MOS for nearly all (179) conditions. This results in an overall correlation of 94.8 %

4.2 Calculation of S-MOS

The relevant parameters for subjects speech quality ratings are level and quality of the transmitted background noise, signal to noise ratio (SNR) between speech and noise in the transmitted signal, improvement or impairment of SNR between non-transmitted and transmitted signal, packet loss, modulation of speech / speech sound and the "naturalness". The principle of the algorithm is shown in figure 3. Besides the change of the SNR, again several parameters based on the Relative Approach and the Δ Relative Approach spectrographs are determined. All parameters are again used as input for a linear quadratic regression.

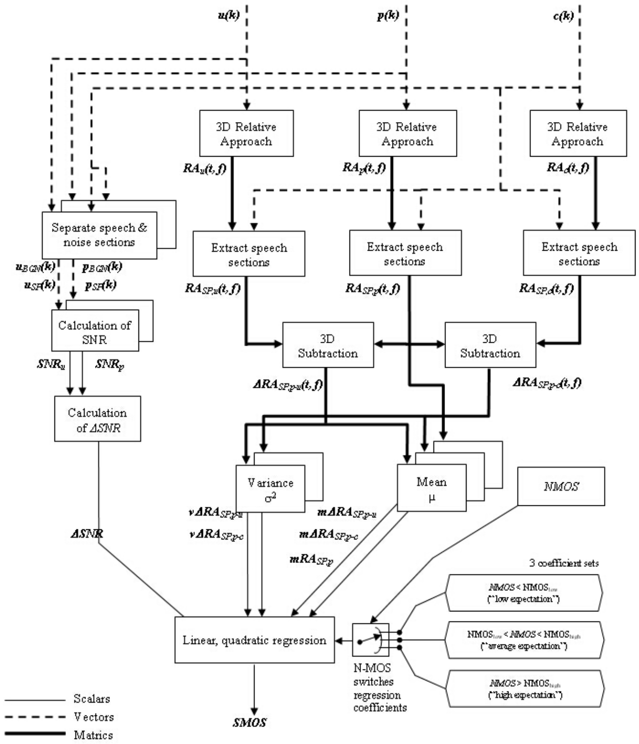

Figure 3: Block diagram of S-MOS calculation

The S-MOS regression depends on the previously calculated N-MOS. The values of the Relative Approach related parameters need to be weighted depending on the N-MOS: if the noise quality is high, also the speech quality is expected to be high (subjects compare the transmitted signal quality to the speech signal without background noise). If the noise quality is low, subjects compare the transmitted speech to the signal of containing speech and background noise. Three groups of N-MOS scores were set up: "low", "average" and "high" speech quality expectation. For each group a coefficient set for the regression was determined [2].

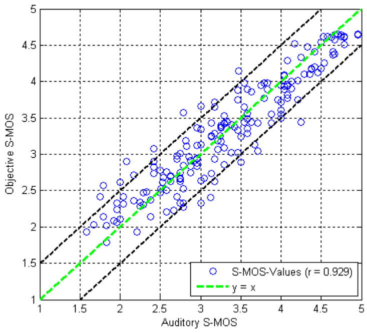

Figure 4: Obj. calculated S-MOS versus audit. S-MOS

Figure 4 shows the objectively calculated S-MOS vs. the auditory S-MOS. The per sample deviation between the subjective and objective S-MOS is less than 0.5 MOS for nearly all (179) conditions. This results in an overall correlation of 92.9 %

4.3 Calculation of G-MOS

The overall speech quality ("global" quality, G-MOS) can be best calculated by using the previously calculated N-MOS and S-MOS as input parameters for a linear quadratic regression. Subjects combine speech and noise quality to a "global" overall quality. The N-MOS and S-MOS algorithm consider all perceptual influences, thus they are the only input parameters for the G-MOS algorithm.

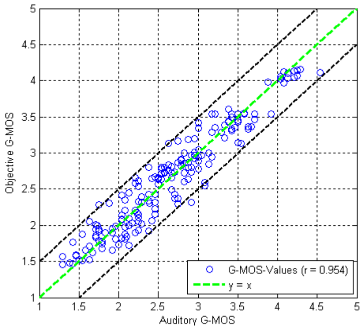

Figure 5: Obj. calculated G-MOS versus audit. G-MOS

Figure 5 shows the objectively calculated G-MOS vs. the auditory G-MOS. The per sample deviation between the subjective and objective G-MOS is less than 0.5 MOS for nearly all (179) conditions. This results in an overall correlation of 95.4 %

5. Conclusion

The new objective model consists of three algorithms which calculate objective MOS values for the speech and noise quality separately (S-MOS, N-MOS) and for the overall quality (G-MOS). The model is applicable for wideband phones and hands-free devices (IP, mobile …) used in realistic background noise environments.

References

[1] ETSI EG 202 396-2: Background Noise Transmission – Network Simulation – Subjective Test Database and Results".

[2] ETSI EG 202 396-3: Background Noise Transmission – Network Simulation – Objective Test Methods

[3] Genuit, K.: Objective Evaluation of Acoustic Quality Based on a Relative Approach, InterNoise '96, Liverpool, UK

[4] Sottek, R.: Modelle zur Signalverarbeitung im menschlichen Gehör, PHD thesis RWTH Aachen, 1993